广电网络安全日志大数据分析的探索与实践

摘　要：为解决日益复杂的网络攻击，通过对广电网络和设备日志的分析，结合云端提供的威胁情报，能够有效发现网络中存在的APT(Advanced Persistent Threat)攻击行为，快速发现网络中受控主机，及时产生告警，实现对海量数据进行多维度快速、自动化的关联分析发现威胁和对异常行为的态势感知，为快速分析、定性、处置提供可视化辅助手段和证据留存。

关键词：网络安全；日志；大数据；APT 攻击
中图分类号：TP393　　**文献标识码：**A
文章编号：1671–0134（2019）06–115–03　　**DOI：**10.19483/j.cnki.11−4653/n.2019.06.035

文 / 朱益中

引言

目前，广电网络的监播监测，传统 DVB 方面已比较到位，新媒体方面也日趋完善。但随着网络应用规模和复杂度的不断提高，网络中传输的数据量急剧上升，网络攻防对抗日趋激烈，企业内部新的安全问题开始显现，复杂的网络环境让安全工作无从下手，攻击者即便是大摇大摆地出入企业的敏感数据区域也无人知晓，投入了大量资金建设的安全防御体系可能成为摆设；传统安全技术对 APT（高级持续性威胁）无能为力，无论是在安全威胁的检测、发现还是响应、溯源等方面都存在严重不足。要解决这些新的安全问题，企业亟须使用新的技术手段来掌控全局的安全态势，从而优化安全运营过程，将企业网络的安全风险控制在合理的区间。《中华人民共和国网络安全法》出台后，对日志留存提出了强制要求，如何监督各地做好网络安全日制留存和收集，并对这些日志开展统一的分析，是企业在新的安全形势下提升安全能力的新契机。

我们的网络和系统在运行的每一个状况信息都使用文字的方式记录下来，称之为日志，可以理解为这是普通人在虚拟世界的行为的记录和投影。日志的类型很多，包括系统日志、应用日志、操作行为日志和数据库日志等，每条日志都记载着时间戳、相关设备名称、使用者及操作行为等相关的描述，将这些日志统一收集起来进行分析，一是可以监控全网日常运行状态，分析问题、追查错误根源、纠正错误；二是能够快速分析整个应用针对最终用户的服务质量；三是分析出安全风险和入侵攻击，及时干预，解除威胁。

网络安全日志大数据分析系统是基于大量网络和系统日志，专用于安全风险和入侵攻击分析的系统，系统主要由数据采集、关联分析、态势感知和可视化呈现四个模块组成：（1）数据采集模块将来自全网几千台设备的日志进行规则化处理，这几千台设备可能是数十个厂商的产品，类型有防火墙、堡垒机、入侵检测等十几种，即使是同一厂商的防火墙，由于生产年代不同，日志格式也可能迥异，要识别这些设备的日志格式，将有效信息入库存储是整个大数据分析系统的基础；（2）关联分析引擎是这个系统的发动机，大数据就如同一座亟待开发的金矿，仅仅存储起来是没有价值的，优秀的分析引擎使用先进的搜索技术，可以迅速感知到异常行为和黑客入侵；（3）态势感知模块是基于威胁情报的辅助手段，将互联网上已知的犯罪手段和特征输入到系统中，让关联分析引擎拥有灵敏的嗅觉；（4）可视化模块要呈现的信息，绝不是简单的 IP 地址和非专业人士看不懂的告警信息，要能够呈现入侵的源和路径，并将关联的资产、部门、人员名称都显示出来。

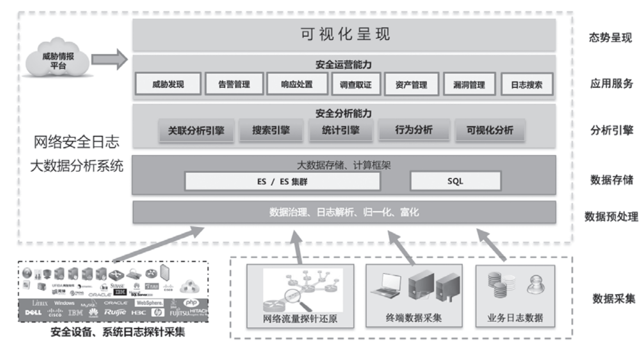


图 1　系统架构设计图

最终我们的目标是希望这个基于威胁情报和日志的系统，能够对安全大数据进行快速、自动化数据分析，全方位监测、发现威胁和异常、快速处置与响应，并针对安全事件进行深入调查。系统通过可视化分析技术将企业总体安全态势进行整体呈现，使安全管理人员和运维人员能够实时掌握安全态势状况，主动发现安全威胁并及时处理，保障业务的顺畅运行。

网络安全日志大数据分析系统的雏形研发完成后，

我们选取了一家在全国广电网络公司中技术实力领先的企业 -- 浙江华数共同开展探索和实践。在实践过程中，通过对各地广电网络公司播出、传输等环节和系统的广泛调研，以及和系统研发人员的深入交流，我们提升了以下三个方面的能力，并解决了一个广电特色的日志采集难题。

1. "全方位、立体化"保证数据分析的全面性

在实践中我们发现，网络的物理出入口并不是攻击者进入到网络的唯一途径，因此，单纯依靠安全设备串行防护以及安全设备日志统计是不足以发现高级风险，从这一点上来说，我们把 PC、智能终端、企业员工、应用软件、企业对外开放的网络服务的全终端采集；从汇聚层交换、核心层交换以及接入层交换的全流量数据采集；从系统级别日志、数据日志、安全设备日志、告警日志的全日志采集。通过三种方式实现"全方位"的能力。

其次，要充分利用云端的安全资源，在新的安全防御体系中，云端的安全资源是整个安全防御的核心，威胁情报和大数据的利用和关联分析，这些数据给未知威胁发现和 APT 攻击检测提供了完全真实网络环境下的大量数据支撑，通过全貌特征"跟踪"攻击者，持续的发现未知威胁，最终确保发现的未知威胁的准确性，进而实现和完善了数据分析的全面性，提高预警发现的能力。

2. 威胁情报数据分类分级和合并管理

威胁情报是基于证据、有关已知或新型威胁或危害的知识，结合公司内部的组织中安全人员不同角色，可提供精准决策参考。但如何实现自动化和设备化处理这些情报，并借助威胁情报数据推动攻击行为分析和溯源查询是个难题。

在系统设计时，首先我们把威胁情报分类分级；其次是将威胁情报数据格式化统一，并且将威胁情报可直接和本地大数据搜索关键词进行匹配和引用，具备上述基础后才可以对网络攻击行为分析和溯源查询，结合搜索技术的数据分析系统可将索引以多个分片和多个副本的形式存储于分布式系统当中，既可提高检索性能，又能保证威胁数据的可靠性和准确性。而且其默认使用的内存索引方式可以保证系统对近期录入的数据做到近乎实时的查询，对于存储于硬盘的 TB 级数据也可做到秒级查询，对全部数据进行数据碰撞和本地风险分析，用来实现基于威胁情报数据推动的攻击行为分析和溯源查询。

3. 采用大数据处理技术来提高系统效率

系统的关联分析核心技术是基于大数据搜索的大数据预警监测分析技术，提升数据查找能力是关键点，在之前传统的方案中，对于本地数据的处理往往采用 SQL 等关系型数据库。这种设计在该项目的当前数据量的处理性能需要下无法满足系统的搜索数据和数据关联。因此，本系统引入大数据搜索技术，将该技术平滑融合到大数据预警监测分析及防御系统中，创新性地采用搜索引擎技术作为本地数据存储和检索核心技术，采用 JSON 格式作为引擎的输入输出格式，在实际测试中发现这样可极大提高检索性能，实现了 TB 级的数据快速搜索能力，同时，相比传统架构也能够降低大量接口上的开发量。

这个技术对于省级广电网络规模的日志的大数据分析挖掘起到了架构基础的作用，从而实现了对大数据快速存储、提取、分析工作，满足了系统在数据处理及分析阶段的实时、高效、并发、可靠的要求，同时也提升了预警监测发现的时间效率，为本地的大规模数据保存、攻击证据留存和查询、实时关联分析提供坚实的技术保障。

这套系统可通过省公司控制中心进行远程配置和权限受理，系统整体部署成本低，具有良好的可扩展性，大大减小各地分公司推广部署的技术难度，避免了传统安全方案需要多次上线的安全风险和硬件成本，还降低了资源消耗以及运维成本。在系统部署过程中，我们还解决了一个广电特色的日志采集难题，广电 DVB 直播网络实际是多个独立小局域网组成的，这些小网的 IP 地址还可能是重复的，如何在不破坏原有的隔离、不改变原来的 IP 地址的情况下，保持各局域网的相对独立又要把日志收集上来，是一个巨大的挑战。

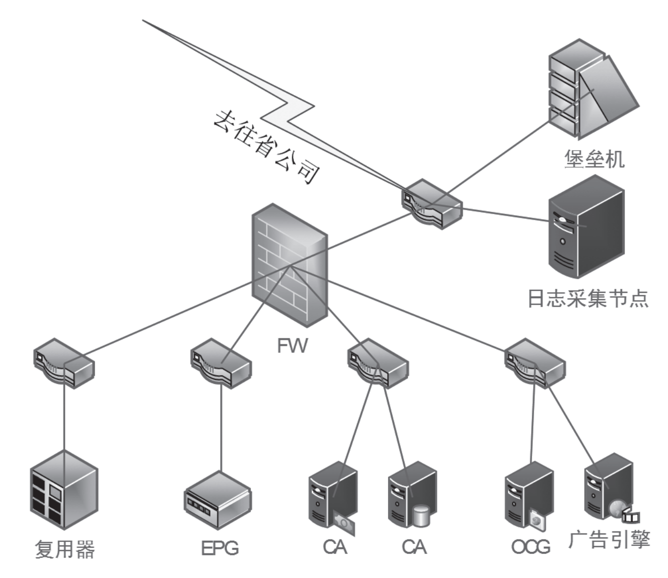


图 2 典型的地市前端部署拓扑图

只有打破围墙式的防御体系，将这些安全孤岛整合起来，打通数据间的隔阂，形成企业或组织的全面数字安全感知体系，才能真正实现安全威胁的积极防御和有效应对。在缺乏总局指导性的操作规范的情况下，我们和华数一起做了大胆的探索，创新的通过防火墙将这些局域网连起来，使得网络之间继续保持隔离，但都能给日志收集服务器传送日志，并将日志的源地址都自动转换为规划好的新地址段。

结语

系统目前已稳定运行将近一年，从实际效果和应用情况来看，原先没有这套系统时，省公司安全工程师在日常安全管理可谓是无所事事，完全看不到各地的情况，只能被动地对安全事件进行响应。系统建成后，提供了一整套“事前预警 + 事中防护 + 事后服务”的全省预警监测分析及防御解决手段，借助横向拓展的大数据能力和分析能力，实现了广泛维度的海量数据采集、处理、展现，并将综合检测的结果与快速响应处置及深入的调查分析进行结合，形成安全事件处置闭环，极大地提升了安全运维的有效性，保障业务顺畅运行。媒

参考文献

[1] 任凯，邓武，俞琰 . 基于大数据技术的网络日志分析系统研究 [J]. 现代电子技术，2016， 39（2）：39–41，44.
[2] 陈世敏 . 大数据分析与高速数据更新 [J]. 计算机研究与发展，2015，52（2）：333–342.
[3]Rossi，Dario Traverso，Stefano Finamore，Alessandro Mellia，Marco Khatouni，Ali Safari Munafo，Maurizio Bocchi，Enrico.Statistical network monitoring: Methodology and application to carrier–grade NAT[J]. Computer networks，2016，107（Oct.9 Pt.1）：20–35.
[4] 李天枫，姚欣，王劲松 . 大规模网络异常流量实时云监测平台研究 [J]. 信息网络安全，2014（9）：1–5
[5]Deka，Ganesh Chandra Walczak，Steven.Special Issue on Bigdata Analytics in Practice[J].Journal of organizational and end user computing，2017，29（4）：vi–viii
[6] 高静，段会 .JSON 数据传输效率研究 [J]. 计算机工程与设计，2011，32（7）：2267–2270.

（**作者单位：**浙江广播电视发展总公司）

（上接第82页）

色通道”，不断提高舆论引导的有效性，严防有害信息及言论的扩散，牢牢把握舆论引导的主动权，为我国改革开放和现代化建设营造良好的舆论氛围。

3.3 有利于实现自主技术创新，推动媒体转型发展

本项目在技术上、功能上和服务上实现全面创新，这必将推动我国大数据产业实现自主技术创新，从而推动我国互联网行业更加健康快速地发展。本项目的建设，对国家信息安全和文化安全、对于抵御西方文化霸权、争夺信息舆论话语权、引导社情民意，具有重大意义。媒

参考文献

[1] 第 43 次《中国互联网络发展状况统计报告》，中国互联网络信息中心，2019（2）：28.
[2] 习近平主持召开网络安全和信息化工作座谈会 [J]. 信息安全与通信保密，2016（5）：10.
[3] 国务院 . 关于在政务公开工作中进一步做好政务舆情回应的通知 [Z].2016–07–30.
[4] 无 . 国务院办公厅关于印发 2018 年政务公开工作要点的通知 [J]. 广东省人民政府公报，2018 （16）：10–15.
[5] 徐茂春 . 网络安全分析中的大数据技术 [J]. 电子技术与软件工程，2018（18）：200.
[6] 董莎莎 . 计算机网络信息安全分析及有效防护措施研究 [J]. 信息与电脑，2018（16）：188–189，192.
[7 张雯文 . 论网络环境的维护与保护 [J]. 法制博览，2019(4)：83–84，82.
[8] 韩卫民，杨柳 .“三农”网络舆情的社会效益与经济效益探析 [J]. 河南农业，2014（21）：15.

（**作者单位：**人民公安报社）